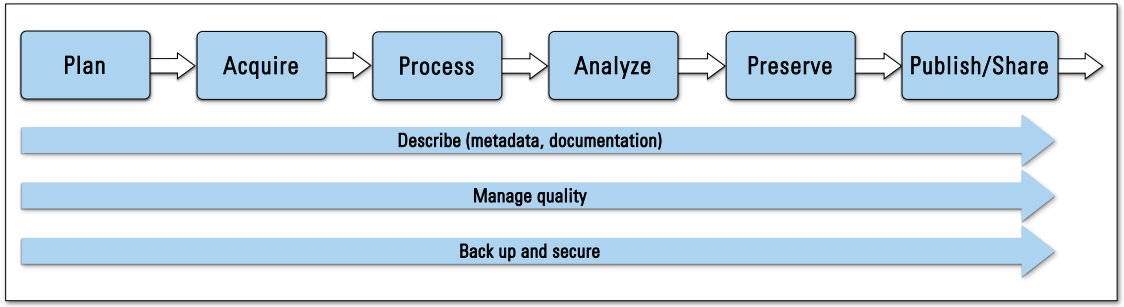
USGS
science for a changing world

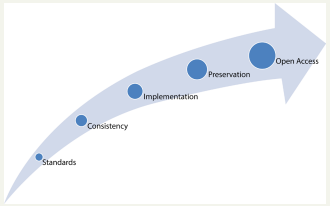
The Evolution and Implementation of the U.S. Geological Survey Science Data Lifecycle Model

John Faundeen, Tom Burley, Jennifer Carlino, David Govoni, Vivian Hutchison, Cassandra Ladino, Elizabeth Martin, Ellyn Montgomery, Steve Tessler, Lisa Zolly



Introduction

The USGS Science Data Lifecycle Model (SDLM) was developed over a span of two years by a USGS Community for Data Integration working group of multi-disciplinary USGS employees. The SDLM provides a high-level view of data—from conception through preservation and sharing—to illustrate how data management (DM) activities relate to project workflows, and to assist with understanding the expectations of proper data management. In applying the Model to research activities, USGS scientists can ensure that data products will be well-described, preserved, discoverable, accessible, and fit for re-use. The Model also serves as a structure to help the USGS evaluate and improve policies and practices for managing scientific data, and to identify areas in which new tools and standards are needed.



Internal Drivers

- Disparate data management practices across programs
- Incomplete implementations of practices
- No shared definitions of data management concepts
- Challenges locating and accessing bureau data assets
- Data not consistently preserved and made accessible for future use

External Drivers

- Open Data Initiative (White House)
- Data.gov and Geoplatform.gov
- Sharing data to address societal challenges

Lifecycle Model Impacts to USGS Data Management

Policy Outcomes: 4 new data management policies (2015)

- Scientific Data Management Foundation
- Metadata for Scientific Data, Software, and Other Information Products
- Review and Approval of Scientific Data for Release
- Preservation Requirements for Digital Scientific Data

Implementation Outcomes: tools, services, and guidance

- USGS Data Management Website (policies, guidance, best practices)
- FAQs (to support implementation of DM Policies)
- USGS Digital Object Identifier Tool (reserve and assign DOIs)
- Metadata Wizard (ArcCatalog plugin) and Online Metadata Editor (web-based tool)
- Science publishing (workflow modifications)
- Data release (systems design and workflows)
- Long-term preservation (policies and procedures)
- USGS Science Data Catalog (discovery of USGS research data)

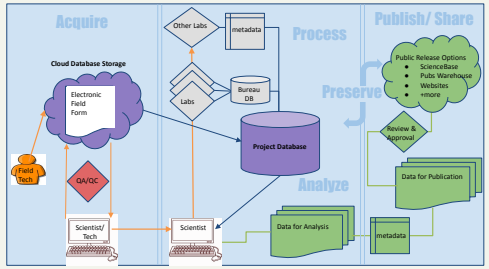
References:

USGS Data Management Website

USGS Open File Report 2013-1265

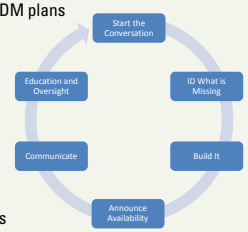
Case Study: USGS Chesapeake Bay Studies

- Data Lifecycle as building block for interdisciplinary DM
  - DM Plan template
  - Data Workflow Model
- Placing project workflows in Data Lifecycle context
  - Scientists relate actions to steps
  - Identify gaps in preserving and documenting data



Case Study: USGS Texas Water Science Center (TXWSC)

- Impetus for operational introspection
- Data Lifecycle used to conduct an internal assessment of scientific DM
  - Project and data lifecycle workflows
  - Existing discipline-specific DM plans
- Staff discussions and survey
  - DM roles & resources
  - Operational needs
- Next steps
  - Develop TXWSC Studies DM Plan, from which projects will inherit baseline requirements
  - Implement requirements into internal project reviews
  - Revise project budgeting to accommodate level of effort
  - Communicate to Project Chiefs the 'why'



U.S. Department of the Interior
U.S. Geological Survey